
Introduction

A very large number of clinical trials have been conducted with human subjects in a wide variety of contexts. Many of these have been concerned, for example, with improving (in some way) the management of patients with disease and others the prevention of the disease or the condition in the first place. The essence of a clinical trial is the comparison of a standard strategy with an alternative (perhaps novel) intervention. The aim of this chapter is to illustrate some of the wide variety of clinical trials that have been conducted and to highlight some key features of their design, conduct and analysis.

1.1 Introduction

The aim of this book is to introduce those who are to become involved with randomized clinical trials to the wide range of challenges that are faced by those who conduct such trials. Our intended readership is therefore expected to range from health care professionals of all disciplines who are concerned with patient care to those more involved with the non-clinical aspects such as the statistical design, data processing and subsequent analysis of the results. We assume no prior knowledge of clinical trial processes and we have attempted to explain the more statistical sections in as non-technical a way as possible. In a first reading of this book, these sections could be omitted. Throughout the book we stress the collaborative nature of clinical trials activity and would hope that readers would consult their more experienced colleagues on aspects of our coverage.

The business of clinical trials is an ongoing process and, as we write, trials are currently being designed, opened, conducted, closed, analyzed and reported. Results are being filtered into current practice and the next trials planned. It is difficult to know where to start in describing the key features of this process, as each stage interacts to some extent with the others. For example, in designing a trial the investigators need to be mindful of the eventual analysis to be undertaken as this governs (but it is only one aspect) of how large a trial should be launched. Some of the steps are intellectually challenging, for example, defining the key therapeutic question, while others may perhaps appear more mundane, such as defining the data forms or the data entry procedures. However, all steps (whether large or small, major or minor) underpin the eventual successful outcome – the influence on clinical practice once the trial results are

available. Entire books have been written for many of these aspects we can only provide an introduction to the process.

Numerous terms need to be introduced, including ‘clinical trial’ itself. As a consequence we have included a Glossary of Terms, which is mainly extracted from Day (2007) *Dictionary of Clinical Trials*. The Glossary defines clinical trial: any systematic study of the effects of a treatment in human subjects. These definitions may not be exhaustive, in the sense that ‘treatment’ used here may be substituted by, for example, ‘intervention’, depending on the specific context of the clinical trial under consideration.

Clinical trials require a multidisciplinary approach in which all partners play a key role at some stage of the trial process. Furthermore, this is the era of evidence-based medicine (EBM), in which it is important to consider critically *all* the available evidence about whether, for example, a treatment works before recommending it for clinical practice. In this respect it is therefore vital that we can clearly see that a proposed trial addresses a key question which will have a clinically meaningful outcome, is well designed, conducted and reported and the results are persuasive enough to change clinical practice if appropriate.

Despite perhaps having a professional interest in the science of clinical trials, everyone has an additional vested interest as potential patients. How many of us have never been to see a doctor, had a hospital admission or taken medication? All of us may be, have been or certainly will be recipients of clinical trial results whether at pre-birth or birth, childhood for vaccination and minor illness, as an adult for fertility, sports injuries, minor and major non-life threatening or life-threatening illnesses and in old age for care related to our mental or physical needs.

1.2 Some completed trials

As we have indicated, there are countless ongoing trials and many have been successfully conducted and reported. To give some indication of the range and diversity of application, we describe a selection of clinical trials that have been conducted. Their designs include some features that we will draw upon in later chapters.

The examples of successfully completed clinical trials illustrate a wide range of topics investigated. These include patients with disease (breast cancer, colon cancer, eczema, glaucoma, malaria and diabetes mellitus), those requiring coronary artery stents or hand surgery, elderly residents of nursing homes, children with dental caries, healthy individuals and those requiring vaccinations. Although not included here, trials are also conducted to evaluate different diagnostic procedures, different bed mattresses to reduce the incidence of bed sores, different dressings for wounds of all types and fertility regulation options for male and females of reproductive potential, for example.

These trials are often termed Phase III trials in contrast to Phase I and Phase II trials, which are concerned with early stages of the (often pharmaceutical) development process. Although the trials differ in aspects of their design, the majority have the general structure of a two (or more) group parallel design in which eligible patients are assigned to receive the alternative options (often treatments but more generally termed interventions) and then at some later time assessed in a way which will be indicative of (successful) outcome. The outcomes measured in these trials include: survival time,

Example 1.1 Recovery of gastrointestinal function after elective colonic resection

Lobo, Bostock, Neal, *et al.* (2002) describe a randomized trial in which 20 patients with colonic cancer either received postoperative intravenous fluids in accordance with current hospital standard practice (*S*) or according to a restricted intake regimen (*R*). A primary endpoint measure in each patient was the solid-phase gastric emptying time on the fourth postoperative day. The observed difference between the median emptying times was shorter with *R* by 56 minutes with 95% confidence interval (CI) from 12 to 132 minutes. The trial also included pre-operative and postoperative (days 1, 2, 4 and 6) measures of the concentrations of serum albumin, haemoglobin and blood urea in a repeated measures design.

Key features include:

- **Design:** randomized comparison of a standard and test, single centre participation, unblinded assessment;
- **Endpoint:** gastric emptying time;
- **Size:** 21 patients following colonic resection;
- **Analysis:** Mann–Whitney-*U* test* for comparing two medians;
- **Conclusion:** The restricted group had shorter delays in returning to gastrointestinal function.

*This can also be referred to as the Wilcoxon Rank-Sum Test.

Example 1.2 Azathioprine for the treatment of atopic eczema

Meggitt, Gray and Reynolds (2006) randomized 63 patients with moderate-to-severe eczema to receive either azathioprine or placebo in a double-blind formulation to ascertain the relative reduction in disease activity determined by the six-area six-sign atopic dermatitis (SASSAD) score between the groups. They reported a 5.4 unit advantage with azathioprine. In this trial patients were randomized, using a minimization procedure, in the ratio of 2 to 1 in favour of azathioprine in order to ‘... encourage recruitment, to reduce the numbers receiving pharmacologically inactive systemic treatment, and to increase the likelihood of identifying infrequent adverse events’.

Key features include:

- **Design:** single centre, randomized double-blind, placebo-controlled, 2 : 1 allocation ratio using minimization;
- **Endpoint:** SASSAD;
- **Size:** 63 patients with moderate-to-severe atopic eczema;
- **Analysis:** comparison of mean group regression slopes over a 12-week period;
- **Conclusion:** azathioprine produces a clinically relevant improvement.

Example 1.3 Anacetrapib and blood pressure

Krishna, Anderson, Bergman, *et al.* (2007) describe a randomized placebo (*P*) controlled, 2-period cross-over trial of anacetrapib (*A*) in 22 healthy volunteers. Half of the individuals were randomized to receive the sequence *AP* (i.e. *A* in Period I of the trial followed by *P* in Period II) and half *PA*. The primary endpoint recorded was the blood pressure on day 10 of Period I and of Period II. The healthy individuals and investigators were blinded to the order in which the trial medication was administered. The authors state: ‘A one-sided test was applied, since another molecule in this class was found to increase blood pressure . . .’. They reported a difference in mean systolic blood pressure between *A* and *P* as 0.6 mm Hg (90% CI -1.54 to 2.74, *p*-value = 0.634) and concluded that: ‘. . ., anacetrapib seems not to increase blood pressure, . . .’.

Key features include:

- **Design:** single centre, randomized placebo controlled, 2-period cross-over trial;
- **Size:** 22 healthy volunteers;
- **Endpoint:** ambulatory blood pressure;
- **Analysis:** comparison of means using analysis of variance;
- **Conclusion:** anacetrapib seems not to increase blood pressure.

Example 1.4 Topical medication and argon laser trabeculoplasty for glaucoma

The Glaucoma Laser Trial Research Group (1995) recruited 271 subjects with newly diagnosed primary-angle glaucoma. One eye of each patient was randomly assigned to argon laser trabeculoplasty (LT) or to a stepped medication (TM) as initial treatment. They treated 261 eyes with LT first followed by TM and the same number with TM first then LT. They found that measures of visual field status for eyes treated by LT-MT were slightly better than those treated by MT-LT. The authors state: ‘Statistical significance was attained for only some of the differences, and the clinical implications of such small differences are not known.’

Key features include:

- **Design:** multicentre, paired design, compares alternative schedules for administering two procedures – the schedule was randomized to one eye with the other eye receiving the alternative;
- **Endpoint:** visual field status;
- **Size:** 271 patients with primary open-angle glaucoma;
- **Analysis:** comparison of means at particular time points following initiation of treatment using the paired *t*-test;
- **Conclusion:** eyes treated with laser trabeculoplasty first were judged to have slightly more improvement and slightly less deterioration.

Example 1.5 Use of glass-ionomer for atraumatic restorative treatment

Lo, Luo, Fan and Wei (2001) conducted a trial in 89 school children from two schools, who had bilateral matched pairs of carious posterior teeth requiring atraumatic restorative treatment (ART). A split-mouth design was used in which the two materials, ChemFlex and Fiji IX GP 49, were randomly placed on contralateral sides. From a total of 101 bilateral matched teeth-pairs included in the trial, the authors concluded that the clinical performance of both materials over a 2-year period was similar.

Key features include:

- **Design:** two schools, split mouth, random allocation;
- **Endpoint:** clinical examination at 24-month recall;
- **Size:** 89 children with 101 pairs of bilateral carious posterior teeth;
- **Analysis:** comparison of mean occlusive wear between materials using a paired *t*-test;
- **Conclusion:** the clinical performance of different materials was similar.

Example 1.6 Use of hip protectors in elderly people in nursing homes

Meyer, Warnke, Bender and Mülhauser (2003) conducted a trial involving 942 residents from 49 nursing homes. In this cluster design, the nursing homes contain 'clusters' of residents and the homes (not the individual residents) were randomized. Twenty-five homes comprising a total of 459 residents were assigned to the intervention group, and 24 homes with 483 residents were assigned to the control group. The intervention comprised a single education session for nursing staff, who then educated residents, and the provision of three hip protectors per resident. The control clusters gave usual care optimized by brief information to nursing staff about hip protectors and the provision of two hip protectors per cluster for demonstration purposes. The main outcome measure was the incidence of hip fractures. There were 21 hip fractures in 21 (4.6%) residents in the intervention group and 42 in 39 (8.1%) residents in the control group – a difference of 3.5% (95% CI 0.3 to 7.3%, *p*-value = 0.072). The authors concluded: 'The introduction of a structured education programme and the provision of free hip protectors in nursing homes may reduce the number of hip fractures'.

Key features include:

- **Design:** multicenter, randomized;
- **Size:** 49 nursing homes comprising 942 residents with high risk of falling;
- **Endpoint:** hip fractures;
- **Analysis:** chi-squared test adjusted for cluster randomization but not for the second fractures in some residents;
- **Conclusion:** increasing the use of hip protectors resulted in a relative reduction of hip fractures of about 40%.

Example 1.7 Treatment of uncomplicated falciparum malaria

Zongo, Dorsey, Rouamba, *et al.* (2007) conducted a randomized non-inferiority trial to test the hypothesis that the risk of recurrent parasitaemia was not significantly worse with artemether-lumefantrine (AL) than with amodiaquine plus sulfadoxine-pyrimethamine (AQ+SP). A total of 826 patients were screened, of which 548 were found to have uncomplicated malaria, and were randomized (273 to AQ+SP and 275 to AL). A primary endpoint was the risk of treatment failure within 28 days of randomization. The authors concluded that AQ+SP with a failure rate of 1.7% (4/233) was more effective than AL with a rate of 10.2% (25/245), representing a difference of 8.5% (95% CI 3 to 12%). These results suggest that the hypothesis of 'non-inferiority' should not be accepted.

Key features include:

- **Design:** multicentre, two-group comparison, non-inferiority trial;
- **Endpoint:** time to recurrent parasitaemia;
- **Size:** large trial of 521 patients with uncomplicated falciparum malaria;
- **Analysis:** comparison of Kaplan–Meier failure-time curves;
- **Conclusion:** AL was less effective than AQ+SP.

Example 1.8 Trastuzumab for HER2-positive breast cancer

Smith, Procter, Gelber, *et al.* (2007) showed that 1 year of treatment with trastuzumab after adjuvant therapy in HER2-positive patients with breast cancer was superior to observation alone. They reported a hazard ratio, $HR = 0.67$ (95% CI 0.47 to 0.91, p -value = 0.0115) for overall survival in favour of adjuvant treatment. This comparison was from two arms of a 3-arm large multicentre international randomized trial comprising 1698 patients randomized to observation alone, 1703 to trastuzumab for 1 year and 1701 to trastuzumab for 2 years: a total of 5102 patients.

Key features include:

- **Design:** randomized, multicentre, observation versus active treatment;
- **Size:** large trial of 5102 women with HER2-positive breast cancer;
- **Endpoint:** overall survival;
- **Analysis:** comparison in 3401 women from the control and 1-year trastuzumab groups using survival curves;
- **Conclusion:** treatment with 1-year trastuzumab after adjuvant chemotherapy has an overall survival benefit.

Example 1.9 Pain prevention following hand surgery

Stevinson, Devaraj, Fountain-Barber, *et al.* (2003) conducted a randomized double-blind, placebo-controlled trial to compare placebo with homeopathic arnica 6C and arnica 30C to determine the degree of pain prevention in patients with carpal tunnel syndrome undergoing elective surgery for their condition. Pain was assessed postoperatively with the short-form McGill Pain Questionnaire (SF-MPQ) at 4 days. A total of 64 patients were randomized to the three groups resulting in median scores of 16.0 (range 0–69), 10.5 (0–76) and 15.0 (0–82) respectively. From these results, the authors suggest that homeopathic arnica has no advantage over placebo in reducing levels of postoperative pain.

Key features include:

- **Design:** single centre, randomized double-blind, placebo-controlled, three-group dose response;
- **Endpoint:** pain using the MPQ;
- **Size:** 64 patients undergoing hand surgery for carpal tunnel syndrome;
- **Analysis:** Kruskal–Wallis test;
- **Conclusion:** irrespective of dose, homeopathic arnica has no advantage over placebo.

Example 1.10 Newly diagnosed patients treated for type 2 diabetes

The randomized trial of Weng, Li, Xu, *et al.* (2008) compared, in newly diagnosed patients treated for type 2 diabetes, three treatments: multiple daily insulin injections (MDI), continuous subcutaneous insulin infusion (CSII) and oral hypoglycaemic agent (OHA).

Key features include:

- **Design:** nine centres, randomized three-group comparison;
- **Endpoint:** time of glycaemic remission;
- **Size:** 410 newly diagnosed patients with type 2 diabetes;
- **Analysis:** Cox-proportional-hazards regression model;
- **Conclusion:** early intensive therapy has favourable outcomes on recovery and maintenance of β -cell function and protracted glycaemic remission compared to OHA.

Example 1.11 Recombinant hepatitis B vaccine

Levie, Gjorup, Skinhøj and Stoffel (2002) compared a 2-dose regimen of recombinant hepatitis B vaccine including the immune stimulant AS04 with the standard 3-dose regimen of Enderix-B in healthy adults. The rationale behind testing a 2-dose regimen was that fewer injections would improve compliance.

Key features include:

- **Design:** two centres, randomized two-group comparison;
- **Endpoint:** seroprotection rate;
- **Size:** 340 healthy subjects aged between 15 and 40 years;
- **Analysis:** Fisher's exact test;
- **Conclusion:** the 2-dose regimen compared favourably to the standard.

Example 1.12 Temporary scaffolding of coronary artery with bio-absorbable magnesium stents

Erbel, Di Mario, Bartunek, *et al.* (2007) describe a non-randomized multicentre trial involving eight centres in which 63 patients were enrolled with single de novo lesions in a native coronary artery. In these patients, a total of 71 biodegradable magnesium stents were successfully implanted. The (composite) primary endpoint was the rate of major adverse cardiac events (MACE) defined as any one of: cardiac death, Q-wave myocardial infarction or target lesion revascularization at 4 months post stent implant. This was to be compared with an anticipated rate of 30%. They reported a rate of MACE of 15/63 (23.8%); all of which were attributed to target lesion revascularization (there were no deaths or Q-wave myocardial infarctions) and concluded: '... stents can achieve an immediate angiographic result similar to ... other metal stents ...'. Nevertheless, the authors also commented in their discussion: 'The absence of randomization precludes direct comparison with other techniques of percutaneous revascularization'.

Key features include:

- **Design:** no comparison group hence non-randomized, multicentre;
- **Size:** 71 stents in 63 patients;
- **Endpoint:** composite endpoint – MACE;
- **Analysis:** proportion experiencing MACE with 95% confidence interval;
- **Conclusion:** bio-absorbable stents can achieve an immediate angiographic result similar to other metal stents and can be safely degraded.

gastric emptying time, reduction in disease activity, visual field status, recurrent parasitaemia, major adverse cardiac events, pain, the number of hip fractures, systolic blood pressure and standard criteria used to assess dental restorations. In the trial of homeopathic arnica for pain relief following hand surgery, assessment was made in a double-blind or double-masked manner in which neither the patient nor the assessor were aware of the treatment received.

The methods used for the allocation to the options included simple randomization of equal numbers per group, a 2 to 1 allocation, a minimization procedure taking into account patient characteristics, randomization to nursing homes (clusters) rather than to individual residents and, in one case, the non-random allocation to a single arm study using a new bio-absorbable stent for coronary scaffolding. In this example, the trial data were compared to that of historical data. For the split-mouth design used for the dental caries trial, a 'random number table was used to determine which tooth of a pair was to be restored with ChemFlex and which with Fiji IX GP'.

The trials ranged in size from 21 patients with colonic cancer to 5102 women with HER2-positive breast cancer. One trial involved 522 eyes from 271 subjects, another 202 teeth from 89 children. Although not fully detailed in the above summaries, methods of statistical analysis ranged from a simple comparison of two proportions to relatively complex methods using techniques for survival time outcomes.

In general, trials are designed to establish a difference between the (therapeutic) options under test, were one to exist. Consequently, they are sometimes termed *superiority* trials. However, in certain circumstances, as in the trial for the treatment of uncomplicated falciparum malaria, the research team were looking for *non-inferiority* implying that the two treatment strategies of AQ + SP and AL would give very similar risks of failure. In the event, the trial suggested that AL was (unacceptably) less effective, implying that non-inferiority was not established. Such designs usually imply that a satisfactory outcome is that the test treatment does not perform worse than the standard to an extent *predefined* by the investigating team. Use of a non-inferiority design often implies that, although some therapeutic loss may be conceded on the main outcome variable, other factors favouring the new therapy will have some features (*gain*) to offset this. For example, if the new compound was a little less effective (not equal to) but had a better toxicity profile, this might be sufficient to prefer it for clinical practice.

1.3 Choice of design

1.3.1 Biological variability

Measurements made on human subjects rarely give exactly the same results from one occasion to the next. Even in adults, our height varies a little during the course of the day. If we measure the blood sugar levels of an individual on one particular day and then again the following day, under exactly the same conditions, greater variation compared to that observed in height would be expected. Hence, were such an individual to be assessed and then receive an intervention (perhaps to lower blood sugar levels), any lowering recorded at the next assessment cannot necessarily be ascribed to the intervention itself. The levels of inherent variability may be very high. This implies that where a subject has an illness,

the oscillations in symptoms may disguise the beneficial effect of the treatment given to improve the condition (at least in the early stages of treatment).

With such variability it follows that, in any comparison made in a biomedical context, differences between subjects or groups of subjects frequently occur. These differences may be due to real effects, random variations or both. It is the job of the experimenter to decide how this variation should be considered in the design of the ensuing trial. Once at the analysis stage, the variation can be suitably partitioned into that due to any real effect of the interventions on the difference between groups, and that from the random or chance component.

Example 1.13 Azathioprine for the treatment of atopic eczema

The considerable between-patient variability in the trial of Example 1.2 is illustrated in Figure 1.1. In the 41 patients receiving azathioprine, the reduction in disease activity (SASSAD) ranged from -10 to 32 . There is considerable overlap of these values with those from the 20 patients receiving placebo, whose values range from -12 to 20 . This figure clearly illustrates that, although there is considerable variation, the majority of patients in both groups improve. Further, the corresponding reduction in percentage body area affected with azathioprine was reported to range from approximately -15 to 85% and for placebo approximately -20 to 45% . Nevertheless, even with the majority of patients improving in both groups, the trial of Meggitt, Gray and Reynolds (2006) indicated a better outcome, on average, for those receiving azathioprine.

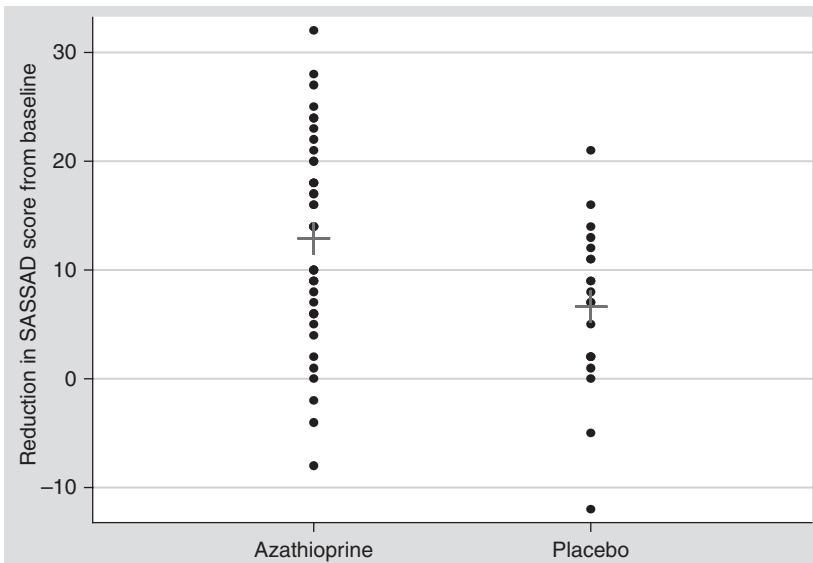


Figure 1.1 Individual patient reductions in disease activity (SASSAD) for the azathioprine and placebo treatment groups with the corresponding means indicated (data from Meggitt, Gray and Reynolds, 2006)

1.3.2 Randomization

In laying the foundations of good experimental design (although more in an agricultural and biological context), Ronald A Fisher (1890–1962) advocated the use of randomization in allocating experimental treatments. For example, in agricultural trials, various plots in a field are randomly assigned to the different experimental interventions. The argument for randomization is that it will prevent systematic differences between the allocated plots receiving the different interventions, whether or not these can be identified by the investigator concerned, before the experimental treatment is applied. Once the experimental treatments are applied and the outcome observed, any differences between treatments can be estimated objectively and without bias. In these and many other contexts, randomization has long been a keystone to good experimental design.

The need for random allocation extends to all experimental situations, including those concerned with patients as opposed to agricultural plots of land. The difficulty arises because clinical trials (less emotive than experiments) do indeed concern human beings who cannot be regarded as experimental units and so allocated the interventions without their consent. The consent process clearly complicates the allocation process and, at least in the past, has been used as a reason to resist the idea of randomization of patients to treatment. Unfortunately the other options, perhaps a comparison of patients receiving a ‘new’ treatment with those from the past receiving the ‘old’, are flawed in the sense that any observed differences (or lack thereof) may not reflect the true situation. In the context of controlled clinical trials, Pocock (1983) concluded, more than 25 years ago and some 30 years after the first randomized trials were conducted, that:

The proper use of randomization guarantees that there is no bias in the selection of patients for the different treatments and so helps considerably to reduce the risk of differences in experimental environment. Randomized allocation is not difficult to implement and enables trial conclusions to be more believable than other forms of treatment allocation.

As a consequence, we focus on randomized controlled trials and do not give much attention to less scientifically rigorous options.

1.3.3 Design hierarchy

The final choice of design for a clinical trial will depend on many factors. The key factors are clearly the specific research question posed, the practicality of recruiting patients to such a design and the resources necessary to support the trial conduct. We shall discuss these and other issues pertinent to the design choice in later chapters. Nevertheless, we can catalogue the main types of design options available; these are listed in Figure 1.2. This gives a relative weight to the evidence obtained from these different types of clinical trial. All other things being equal, the design that maximizes the weight of the resulting evidence should be chosen. For expository purposes, we assume that a comparison of a new test treatment with the current standard for the specific condition in question is being made.

Evidence level	Type of trial
Strongest	Double-blind randomized controlled trial (RCT)
	Single-blind RCT
	Non-blinded (open) RCT
	Non-randomized prospective trial
	Non-randomized retrospective trial
Weakest	Before-and-after design (historical control)
	Case-series

Figure 1.2 The relative strength of evidence obtained from alternative designs for comparative clinical trials

The design that provides the strongest type of evidence is the *double-blind (or double-masked) randomized controlled trial (RCT)*. In this, the patients are allocated to treatment at random. This ensures that, *in the long run*, patients will be comparable in the test and standard groups before treatment commences. Clearly, if the important prognostic factors that influence outcome were known, we could match the patients in the standard and test groups in some way. However, the advantage of randomization is that it balances the *unknown* as well as the *known* prognostic factors, and this could not be achieved by matching. The reason for the attraction of the randomized trial is therefore that it is the *only* design that can give an absolute certainty that there is no bias in favour of one group compared to another at the start of the trial. Indeed, in Example 1.12, Erbel, Di Mario, Bartunek, *et al.* (2007) admitted that failure to conduct a randomized comparison compromised their ability to draw definitive conclusions concerning the stent on test.

For the simple situation in which the attending clinician is also the assessor of the outcome, the trial should ideally be double-blind. This means that neither the patient nor the attending clinician will know the actual treatment allocated. Having no knowledge of which treatment has been taken, neither the patient nor the clinician can be influenced at the assessment stage by such knowledge. In this way, an unprejudiced evaluation of the patient response is obtained. Thus Meggitt, Gray and Reynolds (2006) used double-blind formulations of azathioprine or placebo so that neither the patients with moderate-to-severe eczema, nor their attending clinical team, were aware of who received which treatment. Although they did not give details, the blinding is best broken only at the analysis stage once all the data have been collated.

Despite the inherent advantage of this double-blind design, most clinical trials cannot be conducted in this way as, for example, a means has to be found for delivering the treatment options in an identical way. This may be a possibility if the standard and test are available in tablet form of identical colour, shape, texture, smell and taste. If such ‘identity’ cannot be achieved, then a single-blind design may ensue. In such a design the patient has knowledge of the treatment being given but the clinical assessor does not. In trials with patient survival time as the endpoint, double-blind usually means that both the patient and the treating physician and other staff are blinded; assessment is objective (death) and blinding of the assessor is irrelevant.

Finally, and this is possibly the majority situation, there will be circumstances in which neither the patient nor the assessor can be blind to the treatments actually received. Such designs are referred to as ‘open’ trials.

In certain circumstances, when a new treatment has been proposed for evaluation, all patients are recruited prospectively but allocation to treatment is not made at random. In such cases, the comparisons may well be biased and hence are unreliable. The bias arises because the clinical team choose which patients receive which intervention and in doing so may favour (even subconsciously) giving one treatment to certain patient types and not to others. In addition, the requirement that all patients should be suitable for all options may not be fulfilled; if it is known that a certain option is to be given to a particular subject then we may not rigorously check if the other options are equally appropriate. Similar problems arise if investigators have recruited patients into a single arm study and the results from these patients are then compared with information on similar patients, having (usually in the past) received a relevant standard therapy for the condition in question. However, such historical comparisons are also likely to be biased and to an unknown extent so again it will not be reasonable to ascribe the difference (if any) observed entirely to the treatments themselves. Of course, in either case, there will be situations when one of these designs is the only option available. In such cases, a detailed justification for not using the ‘gold standard’ of the randomized controlled trial is required.

Understandably, in this era of EBM, information from non-randomized comparative studies is categorized as providing weaker evidence than that from randomized trials.

The before-and-after design is one in which, for example, patients are treated with the Standard option for a specified period and then, at some fixed point in time, subsequent patients receive the Test treatment. This is the type of design used by Erbel, Di Mario, Bartunek, *et al.* (2007) to evaluate a bio-absorbable stent for coronary scaffolding. In such examples, the information for the Standard group is retrospective in nature, in that often the information is in the clinical records only and was not initially collected for trial purposes. If this is the case, the before-and-after design is likely to be further compromised. For example, in the ‘before’ period the patient selection criterion, clinical assessments and data recorded may not meet the standards required of the ‘after’ component. Such differences are likely to influence the before-and-after comparison in unforeseen and unknown ways.

Example 1.14 Non-randomized design – glioblastoma in the elderly

Brandes, Vastola, Basso, *et al.* (2003) describe a study comparing radiotherapy alone (Group A), radiotherapy and the combination of procarbazine, lomustine and vincristine (Group B) and radiotherapy with temozolomide (Group C) in 79 elderly patients with glioblastoma. The authors state:

The first group (Group A) was enrolled in the period from March 1993 to August 1995
The second group (group B) was enrolled from September 1995 to September 1997 The
third group (Group C) was enrolled from September 1997 to August 2000 and

The authors conclude:

Overall survival was better in Group C compared with Group A (14.9 months v 11.2 months, $P = 0.002$), but there was no statistical differences found between Groups A and B or between Groups B and C.

However, since patients have not been randomized to groups, we cannot be sure that the differences (and lack of differences) truly reflect the relative efficacy of the three treatments concerned. This type of design should be avoided if at all possible.

A case-series consists of a study in which the experience of an investigator treating a series of patients with a particular approach reports on their outcome. This may be the only ‘design’ option available in rare or unusual circumstances, but is unlikely to provide clear evidence of efficacy. There are many criticisms of this design. Generally, we may not know how the patients have been selected. The clinical team may have an eye for selecting those patients to be given the treatment who are likely to recover in any event. Without further evidence of the natural history of the disease, we do not know whether the patients may have recovered spontaneously without intervention. Finally, we do not know whether their approach to treatment is better than any alternatives.

1.4 Practical constraints

Control of the ‘experiment’ is clearly a desirable feature – perhaps easy to attain in the physics laboratory where experimental conditions are tightly controlled but not so easy with living material, particularly if human. A good trial should answer the questions posed as efficiently as possible. In broad terms, this implies recruiting as few subjects as is reasonably possible for a reliable answer to be obtained.

Although good science may lead to an optimal choice of design, the exigencies of real life may cause these ideals to be modified. We can still keep in mind the hierarchy in the choice of designs of Figure 1.2, but where to enter this hierarchy will depend on circumstance. The investigators therefore do not aim for the best design, but only the best realizable design in their context.

Technical (statistical) aspects of experimental design can be used in a whole variety of settings; nevertheless, there are specific problems associated with implementing these designs in practice in the field of clinical trials. It is clear that trials cannot be conducted without human subjects (often patients); nevertheless, the constraints this imposes are not inconsiderable. Figure 1.3 lists some aspects that need to be considered when conducting such trials.

As we have indicated, the requirements for human studies are usually more stringent than in other research areas. For example safety, in terms of the welfare of the experimental units involved, is of overriding concern in clinical trials but possibly of little relevance in animal studies and of no relevance to laboratory studies. In some sense the laboratory provides, at least in theory, the greatest rigour in terms of the experimental design, and studies in human subjects should be designed (whenever possible) to be as close to these standards as possible. However, no consent procedures from the experimental units or from animals, if they are involved, are required, whereas this is a very important consideration in all human experimentation even in a clinical trial with therapeutic intent.

Constraints may also apply to the choice of interventions to compare. For example, in certain therapeutic trials there may be little chance that a placebo option will bring

Design feature	
Method of assessments	If invasive – may not be acceptable.
Treatment or Intervention	Implicit that treatment should do some good – thus an innocuous or placebo treatment may not be acceptable.
Subject safety issues	Overriding principle is the safety of the subjects
Protocol Review	Scientific and ethical
Consent	Fully informed consent mandatory
Recruitment	Usually, subjects recruited one-by-one over calendar time
Time scale	May be relatively long – rarely weeks, seldom months, quite often years
Trial size	Not too large or too small
Patient losses	Subjects may refuse to continue in the trial at any stage
Observations	Usually, subjects assessed one-by-one over calendar time
Design changes	Almost certainly requires new ethical approval
Data protection	Confidentiality and often National Guidelines for storage and transfer.
Reporting	CONSORT for Phase III trials (Moher, Shultz and Altman, 2001)

Figure 1.3 Special considerations for clinical trials in human subjects

any benefit (although this is certainly not the case in all circumstances). Comparisons may therefore have to be made between two allegedly ‘active’ approaches, despite little direct evidence that either of them will bring benefit. However, if a difference between treatments is demonstrated at the end of such a trial, activity for the better option is established so that comparison with a placebo is not necessary. In contrast, should the two treatments appear not to differ in their effectiveness, no conclusions can be drawn since we do not know whether they are equally beneficial or equally ineffective. An investigating team conducting this type of trial therefore needs to be fully aware of the potential difficulties.

Ethical considerations, as judged perhaps by a local, national or international committee, may also prevent the ‘optimal’ design being implemented. There are also issues related to patient data confidentiality which may, in the circumstances of a multicentre trial, make synthesis of all the trial data problematical. We address other components of Figure 1.3 in later sections of the book.

1.5 Influencing clinical practice

As we have indicated, an important consideration at the design stage of a trial is to consider whether, if the new treatment proves effective, the trial will be reliable enough in itself to convince clinical teams not associated with the trial of the findings. Importantly, if a benefit is established, will this be quickly adopted into national clinical practice? Experience has suggested that all too frequently trials have less impact than they deserve, although it is recognized that results that are adopted in practice are likely to be from trials of an appropriate size, conducted by a respected group with a multi-centre involvement. There are therefore considerations, in some sense outside the strict confines of the design, which investigators should heed if their findings are to have the desired impact.

Some basic or administrative features can help reassure the eventual readers of the reliability of the trial results. These include (some of these may be mandatory) registering the trial itself, involving and informing other clinical colleagues outside the trial team of progress, careful documentation of any serious adverse events, ensuring the trial documentation is complete, establishing procedures for responding to external queries, clarity of the final reporting document in the research literature and seeking avenues for wider dissemination of the trial results.

1.6 History

Probably the single most important contribution to the science of comparative clinical trials was the recognition by Austin Bradford Hill (1897–1991) in the 1940s that patients should be allocated the options under consideration at random, so that comparisons should be free from bias. Consequently, the first randomized trial was planned to test the value of a pertussis vaccine to prevent whooping cough. The results were subsequently published by the Medical Research Council Whooping-Cough Immunization Committee (1951). He later stated: ‘The aim of the controlled clinical trial is very simple: it is to ensure that the comparisons we make are as precise, as informative and as convincing as possible.’ This development by itself may not have led directly to more theoretically based statistical innovation, but was the foundation for the science of clinical trials.

Nevertheless, the history of clinical trials research precedes this important development by many years. Clinical trials were mentioned by Avicenna (980–1037) in *The Canon of Medicine* (1025), in which he laid down rules for the experimental use and testing of drugs and wrote a precise guide for practical experimentation in the process of discovering and proving the effectiveness of medical drugs and substances. His rules and principles for testing the effectiveness of new drugs and medications are summarized in Figure 1.4, and still form the basis of modern clinical trials.

- (1) The drug must be free from any extraneous accidental quality.
- (2) It must be used on a simple, not a composite, disease.
- (3) The drug must be tested with two contrary types of diseases, because sometimes a drug cures one disease by its essential qualities and another by its accidental ones.
- (4) The quality of the drug must correspond to the strength of the disease. For example, there are some drugs whose heat is less than the coldness of certain diseases, so that they would have no effect on them.
- (5) The time of action must be observed, so that essence and accident are not confused.
- (6) The effect of the drug must be seen to occur constantly or in many cases, for if this did not happen, it was an accidental effect.
- (7) The experimentation must be done with the human body, for testing a drug on a lion or a horse might not prove anything about its effect on man.

Figure 1.4 Avicenna’s rules for the experimental use and testing of drugs

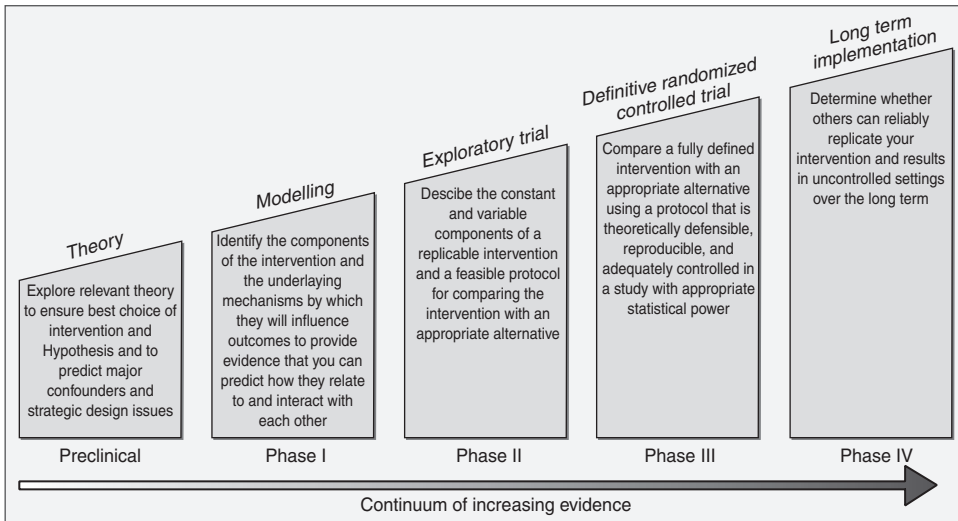


Figure 1.5 Sequential phases of developing randomized controlled trials of complex interventions (from Campbell, Fitzpatrick, Haines, *et al.*, 2000)

One of the most famous clinical trials was that conducted by James Lind (1716–1794) in 1747. He compared the effects of various different acidic substances, ranging from vinegar to cider, on groups of sailors afflicted with scurvy, and found that the group who were given oranges and lemons had largely recovered from their scurvy after 6 days. Somewhat later, Frederick Akbar Mahomed (1849–1884) founded the Collective Investigation Record for the British Medical Association. This organization collated data from physicians practicing outside the hospital setting and was an important precursor of modern collaborative clinical trials.

The very nature of clinical trials research is multidisciplinary in nature so that a team effort is always needed from the concept stage through design, conduct, monitoring and reporting. This collaborative effort has not only led to medical developments in many areas but also to developments of a more statistical nature. For those working in cancer and for whom survival was a key endpoint in the clinical trials, the two seminal papers published by Peto, Pike, Armitage, *et al.* (1976, 1977) in the *British Journal of Cancer* marked a new era. These papers provided the template for key items essential to the design, conduct, analysis and reporting of randomized trials, with emphasis on those requiring prolonged observation of each patient. In particular, these papers described the Kaplan and Meier (1958) estimate of the survival curve, logrank test and the stratified logrank test in such detail that any careful investigator could follow the necessary steps. A computer program (termed the Oxford program) had also been distributed (some time before the date of the publications themselves) and this allowed the methods suggested by the papers to be implemented. Certainly, for those working in data centres with responsibility for many (often reasonably large) trials, this program facilitated the analysis and helped to ensure that the ideas expressed in these articles were widely disseminated. These papers formed the basic text for those involved in clinical trials

for many years and, as well as making the ideas accessible to medical statisticians, their role in easing the acceptance of statistical ideas into the clinical community cannot be underestimated.

It should not go unnoticed that DR Cox was one of the authors of the seminal papers referred to above, although his paper describing the proportional hazards regression model appeared some 4 years earlier (Cox, 1972). His paper was presented at a discussion meeting of the UK Royal Statistical Society and subsequently published in Series B of the Society's journals. This journal deals with the more theoretical aspects of statistical research; it does not make easy reading for many statisticians and would not be one to which clinical teams might readily refer. Despite this, this particular paper is probably one of the most cited papers in the medical literature. The methodology leads to easier analysis of trials, with survival time endpoints that include stratification in their design and/or baseline patient characteristics at the time of randomization which may affect prognosis.

As we have indicated, EBM requires that it is important to critically assess *all* the available evidence about whether an intervention works. More recently, systematic overviews have become a vital component of clinical trial research. They are routinely applied *before* launching new trials, as a means of confirming the need to carry out a clinical trial, and *after* completing trials, as a means of synthesizing and summarizing the current knowledge on the topic of interest. These reviews are the focal interest of the Cochrane Collaboration; the associated handbook by Higgins and Green (2005) provides the key to their implementation.

Some developments have not depended on technical advancement, such as the now standard practice of reporting confidence intervals rather than relying solely on *p*-values at the interpretation stage. Over this same time period the expansion in data processing capabilities and the range of analytical possibilities, made possible by the amazing development in computer power, have been of major importance. Despite many advances, the majority of randomized controlled trials remain simple in design – most often a two-group comparison.

1.7 How trials arise

Although the focus of this book is on comparative or Phase III trials to establish the relative efficacy of the interventions under test, it should be recognized that these may be preceded by an often extensive research programme. This programme may start with the laboratory bench, moving to animal studies and then to early and later stage studies in man. Also, once the Phase III stage itself is complete, there may be further studies initiated. Figure 1.5, taken from Campbell, Fitzpatrick, Haines, *et al.* (2000), succinctly summarizes the pathway of the whole trial process.

The steps range from studies to determine the pharmacokinetic profile of a drug in healthy volunteers (Preclinical) to establishing the appropriate dosage for use in man (Phase I), then the establishment of indications of activity (Phase II). However, some of these steps may be taken in parallel and even simultaneously in the same subjects.

These early studies are not usually randomized. However, the studies conducted by Krishna, Anderson, Bergman, *et al.* (2007) are described as ‘randomized’ and ‘phase I’. Randomized they undoubtedly are, but their use of the Phase I nomenclature is not compatible with Figure 1.5. This highlights a difficulty when attempting to categorize trials using such a simple system. We may imagine that there will be clear stages in the development of a bio-absorbable coronary stent. These too will not exactly parallel those of drug development, although they may well involve laboratory and animal studies. Thus the single arm trial of Erbel, Di Mario, Bartunek, *et al.* (2007) may be considered as close to the Phase II type.

There are also parallels (although modifications will be necessary) for new approaches to, for example, surgical, radiotherapy or physiotherapy techniques and combinations of different procedures. They also extend beyond merely therapeutic trials to planning, for example, trials comparing alternative forms of contraception in women, and those evaluating alternative health promotion interventions. However in some instances, such as in trials comparing educational packages, they may start at the full Phase III stage without involving the earlier phases.

Alternatively, comparative trials may evolve from questions arising in clinical practice and not from a specific development process. We may therefore wish to compare different surgical timings, at 6 months or at 1 year of age, for reconstructive surgery in infants with cleft palate as is proposed in the trial being conducted by Yeow, Lee, Cheng, *et al.* (2007).

Whatever the pathway, the eventual randomized comparative trial to be conducted is clearly a major event. Only when this has been conducted will there be reliable (although not necessarily convincing) evidence of the efficacy of the intervention concerned. In certain situations, often for regulatory purposes, a Phase III trial may be followed by a confirmatory trial asking essentially the same question. In addition, following the regulatory approval of a product, so-called Phase IV or post-marketing trials may be initiated with the aim of gaining broader experience in using the new product.

1.8 Ethical considerations

For a trial to be ethical, at the time it is designed the ethical review committees will want to be convinced that there is collective uncertainty among clinicians as to which treatment is superior or more appropriate for the patients. They will also need to be persuaded that the sample size and other aspects of the study design are such that the trial is likely to provide information sufficient to reduce this uncertainty and therefore influence subsequent medical practice if one treatment or the other appears superior.

A clinical trial cannot go forward until the protocol has been through the appropriate ethical review processes, the exact nature of which varies from country to country. These should always include a very thorough review of the scientific aims as well as the more subject-oriented concerns to protect those who will be recruited to the trial. Briefly, this implies if a trial is not scientifically sound then it should not be judged as ethically acceptable.

1.9 Regulatory requirements

In addition to the more overtly scientific parts of the clinical trials process on which to focus, there are many regulatory requirements which a trial team are obliged to adhere to. For example, the regulations insist that informed consent is obtained from patients entering trials and on the preservation of personal data confidentiality. These regulations are generally referred to as requirements for Good Clinical Practice (GCP) as is described in ICH (1996). We will refer to specific aspects of GCP as they arise in the text, but readers are cautioned that the specifics are continually being changed. Principles to guide statisticians working on clinical trials have been laid down by ICH E9 (1998) and ICH E9 Expert Working Group (1999).

If the trial is seeking regulatory approval of (say) a new drug, then all the associated requirements for approval should be reviewed by the trial team *before*, *during* and *after* the development of the trial protocol to avoid the rejection of the application on what might be a technical detail. For example, there may be a regulatory requirement for some additional animal studies to be conducted before approval can be granted. These requirements are summarized in documents such as those of US Food and Drug Administration (FDA, 1988) and European Medicines Agency (EMA, 2009).

In some circumstances, it is a requirement for regulatory approval that a confirmatory trial is conducted. Such a trial is essentially a repeat of an initial trial, perhaps in a different or wider patient group or with wider clinical teams involvement, but it must follow the essential features of the predecessor design. Clearly these details should be cross checked with the relevant authorities before the protocol is finalized and patients are recruited.

1.10 Focus

As we have illustrated, the size of clinical trials can range from the relatively few to as many as several thousands of subjects being recruited. Consequently, and leaving specific details aside, these will require a range of resources from the relatively modest to the very considerable. It must be emphasized that the size of a clinical trial is determined by the question(s) that are posed, and the resources allocated should reflect the importance of that question. Clearly a very experienced team is required to launch a large trial, but even the design team of an ultimately small-sized trial will need access to appropriate personnel including, at a minimum, those with clinical, statistical, data management and organizational skills as well as other specialist skills such as pharmacy or pathology. It is important that the design team do not underestimate the scale of the task.

The focus of this book is on the design of (randomized) comparative (usually termed Phase III) trials which are likely to be of a relatively modest size. We aim to provide clear guidance as to how these may be designed, conducted, (to some extent) analyzed and reported. However, it is also important that investigators contributing patients to clinical trials who are perhaps not part of the design team also understand the issues concerned; the very success of the trials depends crucially on their collaboration and understanding of the processes involved.

1.11 Further reading

Although Day (2007) provides a comprehensive list of books about clinical trials the following are particularly useful:

Day, S. (2007) *Dictionary for Clinical Trials*, 2nd edn, John Wiley & Sons, Ltd, Chichester.

Fitzpatrick, S. (2008a) *Clinical Trial Design*, ICR Publishing, Marlow, www.icr-global.org

Fitzpatrick, S. (2008b) *The Clinical Trial Protocol*, ICR Publishing, Marlow, www.icr-global.org

Girling, D.J., Parmar, M.K.B., Stenning, S.P., . Stephens, R.J., and Stewart, L.A.,(2003) *Clinical Trials in Cancer*, Oxford University Press, Oxford.

Institute of Clinical Research (2008) *The Fundamental Guidelines for Clinical Research V2.0*, ICR Publishing, Marlow.

Jadad, A. (1998) *Randomised Controlled Trials*, British Medical Journal, London.

Machin, D., Campbell, M.J., Tan, S.B. and Tan, S.H. (2009) *Sample Size Tables for Clinical Studies*, 3rd edn, Wiley-Blackwell, Chichester.

Redwood, C. and Colton, T. (eds) (2001) *Biostatistics in Clinical Trials*, John Wiley & Sons, Ltd, Chichester.

Senn, S. (2007) *Statistical Issues in Drug Development*, 2nd edn, John Wiley & Sons, Ltd, Chichester.

Wang, D. and Bakhai, A. (eds) (2006) *Clinical Trials, A Practical Guide to Design, Analysis and Reporting*, Remedica, London.

Hints on how to display medical data in tabular and graphical form are given by:

Freeman, J.V., Walters, S.J. and Campbell, M.J. (2008) *How to Display Data*, BMJ Books, Oxford.

For those specifically interested in health related quality of life issues:

Fayers, P.M. and Machin, D. (2007) *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*, John Wiley & Sons, Ltd, Chichester.

For those requiring a wide view of how randomised trials have impacted on clinical practice over a wide range of diseases and conditions:

Machin, D., Day, S. and Green, S. (eds) (2006) *Textbook of Clinical Trials*, 2nd edn, John Wiley & Sons, Ltd, Chichester.

